Large Deviations and Sum Rules for Orthogonal Polynomials

CLAPEM XIV

Universidad de Costa Rica, December, 2016

Barry Simon

IBM Professor of Mathematics and Theoretical Physics, Emeritus

California Institute of Technology

Pasadena, CA, U.S.A.

Lecture 3: The Theory of Large Deviations

Rate Functions
Cramér's Theorem
Contraction Principle
Projective Limits
Killip–Nenciu Theorem

- Lecture 1: OPRL, OPUC and Sum Rules
- Lecture 2: Meromorphic Herglotz Functions and Proof of KS Sum Rule
- Lecture 3: The Theory of Large Deviations
- Lecture 4: GNR Proof of Sum Rules

References

[DS] J. Deuschel and D. Stroock, *Large Deviations*, Academic Press, Boston, 1989.

[DZ] A. Dembo and O. Zeitouni *Large Deviations and Applications*, Springer, Berlin, 1998.

[KN] R. Killip and I. Nenciu, *Matrix models for circular ensembles*, Int. Math. Res. Not., **50** (2004), 2665–2701.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

We consider a sequence of probability measures, $\{\mathbb{P}_n\}_{n=1}^{\infty}$, on a space, X.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

We consider a sequence of probability measures, $\{\mathbb{P}_n\}_{n=1}^{\infty}$, on a space, X. In some of our applications, X will be a space of measures, so we'll need to avoid getting confused about measures on measures!

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

We consider a sequence of probability measures, $\{\mathbb{P}_n\}_{n=1}^{\infty}$, on a space, X. In some of our applications, X will be a space of measures, so we'll need to avoid getting confused about measures on measures! Naively, one has a Large Deviation Principle (aka LDP) if the $\mathbb{P}_n$–probability that x is near x_0 is $\mathrm{O}(e^{-nI(x_0)})$.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

We consider a sequence of probability measures, $\{\mathbb{P}_n\}_{n=1}^\infty$, on a space, X. In some of our applications, X will be a space of measures, so we'll need to avoid getting confused about measures on measures! Naively, one has a Large Deviation Principle (aka LDP) if the $\mathbb{P}_n$–probability that x is near x_0 is $\mathrm{O}(e^{-nI(x_0)})$. I is called the *rate function*.

Intuitive Ideas

In this lecture, I'll sketch some of the key ideas in the theory of large deviations. Two books on the subject are Deuschel–Stroock and Dembo–Zeitouni. As both books say, there is not so much a theory as a set of powerful tools. While the subject goes back to Laplace, the modern framework is due to Varadhan, Donsker–Varadhan and Freidlin–Wentzel in the 1960's and 1970's. Varadhan got the Abel Prize for his work on this subject.

We consider a sequence of probability measures, $\{\mathbb{P}_n\}_{n=1}^{\infty}$, on a space, X. In some of our applications, X will be a space of measures, so we'll need to avoid getting confused about measures on measures! Naively, one has a Large Deviation Principle (aka LDP) if the $\mathbb{P}_n$–probability that x is near x_0 is $\mathrm{O}(e^{-nI(x_0)})$. I is called the *rate function*. We allow replacing n in the exponent by a_n called the *speed*.

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book).

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book). Let I be a non–negative function on X and $\{a_N\}_{N=1}^{\infty}$ a sequence of positive numbers with $a_N \to \infty$.

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^\infty$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book). Let I be a non–negative function on X and $\{a_N\}_{N=1}^\infty$ a sequence of positive numbers with $a_N \to \infty$. We say that $\{\mathbb{P}_N\}_{N=1}^\infty$ obeys a LDP with rate function, $I : X \to [0, \infty]$, and speed $\{a_N\}_{N=1}^\infty$ if and only if

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book). Let I be a non–negative function on X and $\{a_N\}_{N=1}^{\infty}$ a sequence of positive numbers with $a_N \to \infty$. We say that $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP with rate function, $I : X \to [0, \infty]$, and speed $\{a_N\}_{N=1}^{\infty}$ if and only if

1. I is non-negative and lower semicontinuous on X

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book). Let I be a non–negative function on X and $\{a_N\}_{N=1}^{\infty}$ a sequence of positive numbers with $a_N \to \infty$. We say that $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP with rate function, $I : X \to [0, \infty]$, and speed $\{a_N\}_{N=1}^{\infty}$ if and only if

1. I is non-negative and lower semicontinuous on X
2. For every open set, $U \subset X$, we have that

$$\liminf_{N\to\infty} \frac{1}{a_N} \log \mathbb{P}_N(U) \geq -\inf_{x\in U} I(x)$$

Precise Definition

Let $\{\mathbb{P}_N\}_{N=1}^\infty$ be a sequence of probability measures on a Polish space, X (complete metric space; the theory of measures on Polish spaces is discussed in Section 4.14 of my *Real Analysis* book). Let I be a non–negative function on X and $\{a_N\}_{N=1}^\infty$ a sequence of positive numbers with $a_N \to \infty$. We say that $\{\mathbb{P}_N\}_{N=1}^\infty$ obeys a LDP with rate function, $I : X \to [0, \infty]$, and speed $\{a_N\}_{N=1}^\infty$ if and only if

1. I is non-negative and lower semicontinuous on X
2. For every open set, $U \subset X$, we have that
$$\liminf_{N\to\infty} \frac{1}{a_N} \log \mathbb{P}_N(U) \geq - \inf_{x\in U} I(x)$$
3. For every closed set, $K \subset X$, we have that
$$\limsup_{N\to\infty} \frac{1}{a_N} \log \mathbb{P}_N(K) \leq - \inf_{x\in K} I(x)$$

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion.

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion. A rate function is called *good* if for each positive α, $\{x \,|\, I(x) \leq \alpha\}$ is compact (note that by the lower semicontinuity, this set is always closed).

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion. A rate function is called *good* if for each positive α, $\{x \mid I(x) \leq \alpha\}$ is compact (note that by the lower semicontinuity, this set is always closed).

We pause to hint at the key idea of Gamboa, Nagel and Rouault (GNR). Suppose we have a sequence of probability measures on the set of probability measure on $\partial\mathbb{D}$.

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion. A rate function is called *good* if for each positive α, $\{x \mid I(x) \leq \alpha\}$ is compact (note that by the lower semicontinuity, this set is always closed).

We pause to hint at the key idea of Gamboa, Nagel and Rouault (GNR). Suppose we have a sequence of probability measures on the set of probability measure on $\partial\mathbb{D}$. Since the Verblunsky map is homeomorphism, it induces a family of probability measures on the set of Verblunsky coefficients.

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion. A rate function is called *good* if for each positive α, $\{x \mid I(x) \leq \alpha\}$ is compact (note that by the lower semicontinuity, this set is always closed).

We pause to hint at the key idea of Gamboa, Nagel and Rouault (GNR). Suppose we have a sequence of probability measures on the set of probability measure on $\partial\mathbb{D}$. Since the Verblunsky map is homeomorphism, it induces a family of probability measures on the set of Verblunsky coefficients. Clearly an LDP on one side implies one on the other.

Precise Definition

We note that the rate function is uniquely determined by these conditions because lower semicontinuity implies that $I(x_0) = \lim_B \inf_{x \in B} I(x)$ where the limit is over the directed set of all open (or over all closed) neighborhoods of x_0, directed by inverse inclusion. A rate function is called *good* if for each positive α, $\{x \mid I(x) \leq \alpha\}$ is compact (note that by the lower semicontinuity, this set is always closed).

We pause to hint at the key idea of Gamboa, Nagel and Rouault (GNR). Suppose we have a sequence of probability measures on the set of probability measure on $\partial\mathbb{D}$. Since the Verblunsky map is homeomorphism, it induces a family of probability measures on the set of Verblunsky coefficients. Clearly an LDP on one side implies one on the other. If we can independently compute the rate function on each side, we get an equality of positive functions, i.e. a positive sum rule!!!

Two Hilfssatz

The following two elementary results will be useful.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x\to\partial U\cup\{\infty\}} G(x) = \infty$ and $\inf_{x\in U} G(x) = 0$.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x \to \partial U \cup \{\infty\}} G(x) = \infty$ and $\inf_{x \in U} G(x) = 0$. Let $F \in L^1(\mathbb{R}^\nu)$ with $\operatorname{supp} F \subset \overline{U}, F \geq 0$ and $\inf_{x \in K} F(x) > 0$ for all compact $K \subset U$.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x\to\partial U\cup\{\infty\}} G(x) = \infty$ and $\inf_{x\in U} G(x) = 0$. Let $F \in L^1(\mathbb{R}^\nu)$ with $\text{supp}F \subset \overline{U}, F \geq 0$ and $\inf_{x\in K} F(x) > 0$ for all compact $K \subset U$. Let $d\mathbb{P}_N(x) = Z_N^{-1} e^{-NG(x)} F(x) d^\nu x$ where $Z_N = \int e^{-NG(x)} F(x) d^\nu x$.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x\to\partial U\cup\{\infty\}} G(x) = \infty$ and $\inf_{x\in U} G(x) = 0$. Let $F \in L^1(\mathbb{R}^\nu)$ with $\mathrm{supp} F \subset \overline{U}, F \geq 0$ and $\inf_{x\in K} F(x) > 0$ for all compact $K \subset U$. Let $d\mathbb{P}_N(x) = Z_N^{-1} e^{-NG(x)} F(x) d^\nu x$ where $Z_N = \int e^{-NG(x)} F(x) d^\nu x$. Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function G.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x\to\partial U\cup\{\infty\}} G(x) = \infty$ and $\inf_{x\in U} G(x) = 0$. Let $F \in L^1(\mathbb{R}^\nu)$ with $\text{supp} F \subset \overline{U}, F \geq 0$ and $\inf_{x\in K} F(x) > 0$ for all compact $K \subset U$. Let $d\mathbb{P}_N(x) = Z_N^{-1} e^{-NG(x)} F(x) d^\nu x$ where $Z_N = \int e^{-NG(x)} F(x) d^\nu x$. Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function G.

Remarks 1. The assumptions in the second are overly strong but suffice for the applications we make below.

Two Hilfssatz

The following two elementary results will be useful.

1. Let $a_N = N^\ell$ (for $\ell > 0$). Fix N_0. Then $\{\mathbb{P}_{N+N_0}\}_{N=1}^\infty$ obeys a LDP with speed a_N and rate function, I, if and only if $\{\mathbb{P}_N\}_{N=1}^\infty$ does
2. Let $U \subset \mathbb{R}^\nu$ be open. Let G be continuous on U with $\lim_{x\to\partial U\cup\{\infty\}} G(x) = \infty$ and $\inf_{x\in U} G(x) = 0$. Let $F \in L^1(\mathbb{R}^\nu)$ with $\operatorname{supp} F \subset \overline{U}, F \geq 0$ and $\inf_{x\in K} F(x) > 0$ for all compact $K \subset U$. Let $d\mathbb{P}_N(x) = Z_N^{-1} e^{-NG(x)} F(x) d^\nu x$ where $Z_N = \int e^{-NG(x)} F(x) d^\nu x$. Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function G.

Remarks 1. The assumptions in the second are overly strong but suffice for the applications we make below.

2. The proofs are straight–forward.

Statement of Cramér's Theorem

Given a random variable, ξ, let

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda \in \mathbb{R}} \left(\lambda x - \Lambda(\lambda)\right)$$

its Legendre transform.

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ.

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda\in\mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ. **Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function I.**

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ. **Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function I.** This is known as Cramér's Theorem.

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ. **Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function I.** This is known as Cramér's Theorem.

Remark. By Jensen's inequality, Λ is convex and as a $\sup$ of linear functions, so is $I(x)$.

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ. **Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function I.** This is known as Cramér's Theorem.

Remark. By Jensen's inequality, Λ is convex and as a $\sup$ of linear functions, so is $I(x)$. If Λ is everywhere finite, it is C^1, $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$ and $I(x) > 0$ for $x \neq \bar{x}$.

Statement of Cramér's Theorem

Given a random variable, ξ, let

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda\xi})$$

be its cumulant generating function and

$$I(x) = \sup_{\lambda\in\mathbb{R}} (\lambda x - \Lambda(\lambda))$$

its Legendre transform. Let $\mathbb{P}_N$ be the probability distribution for $N^{-1}S_N \equiv N^{-1}(X_1 + \cdots + X_N)$, where $\{X_j\}_{j=1}^{\infty}$ are independent copies of ξ. **Then $\mathbb{P}_N$ obeys a LDP with speed N and good rate function I.** This is known as Cramér's Theorem.

Remark. By Jensen's inequality, Λ is convex and as a $\sup$ of linear functions, so is $I(x)$. If Λ is everywhere finite, it is C^1, $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$ and $I(x) > 0$ for $x \neq \bar{x}$. Thus this result amplifies the law of large numbers.

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line.

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex,

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex, with $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$, and further that

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex, with $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$, and further that

$$I(x) = x\lambda_x - \Lambda(\lambda_x) \,\text{with}\, \Lambda'(\lambda_x) = x \,\text{and}\, \lambda_x \geq 0 \text{ iff } x \geq \bar{x}$$

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex, with $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$, and further that

$$I(x) = x\lambda_x - \Lambda(\lambda_x) \,\text{with}\, \Lambda'(\lambda_x) = x \,\text{and}\, \lambda_x \geq 0 \text{ iff } x \geq \bar{x}$$

The key point is that a solution to the equation $\Lambda'(\lambda_x) = x$ exists;

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex, with $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$, and further that

$$I(x) = x\lambda_x - \Lambda(\lambda_x) \,\text{with}\, \Lambda'(\lambda_x) = x \,\text{and}\, \lambda_x \geq 0 \text{ iff } x \geq \bar{x}$$

The key point is that a solution to the equation $\Lambda'(\lambda_x) = x$ exists; it is mostly here that we use the assumptions that the image of ξ is $\mathbb{R}$ and that $\mathrm{Dom}(\Lambda) = \mathbb{R}$.

Proof of Cramér's Theorem

We sketch the proof in case $\Lambda(\lambda) < \infty$ for all λ and the image of ξ is the whole real line. Under these conditions, a bit of calculus and convex analysis shows that $I(x)$ is non-negative, strictly convex, with $I(\bar{x}) = 0$ where $\bar{x} = \mathbb{E}(\xi)$, and further that

$$I(x) = x\lambda_x - \Lambda(\lambda_x) \,\text{with}\, \Lambda'(\lambda_x) = x \,\text{and}\, \lambda_x \geq 0 \text{ iff } x \geq \bar{x}$$

The key point is that a solution to the equation $\Lambda'(\lambda_x) = x$ exists; it is mostly here that we use the assumptions that the image of ξ is $\mathbb{R}$ and that $\mathrm{Dom}(\Lambda) = \mathbb{R}$.

We start with the proof of the upper bound on probabilities.

Proof of Cramér's Theorem

To see the large deviations upper bound, we note first that the strict convexity of $I(\cdot)$ implies that the latter is strictly monotone increasing (resp. decreasing) on $[\bar{x}, \infty)$ (resp. $(-\infty, \bar{x}]$).

Proof of Cramér's Theorem

To see the large deviations upper bound, we note first that the strict convexity of $I(\cdot)$ implies that the latter is strictly monotone increasing (resp. decreasing) on $[\bar{x}, \infty)$ (resp. $(-\infty, \bar{x}]$)). It is therefore enough to prove the upper bound on intervals of the form $(-\infty, x]$ (with $x < \bar{x}$) or $[x, \infty)$ (with $x > \bar{x}$).

Proof of Cramér's Theorem

To see the large deviations upper bound, we note first that the strict convexity of $I(\cdot)$ implies that the latter is strictly monotone increasing (resp. decreasing) on $[\bar{x}, \infty)$ (resp. $(-\infty, \bar{x}]$)). It is therefore enough to prove the upper bound on intervals of the form $(-\infty, x]$ (with $x < \bar{x}$) or $[x, \infty)$ (with $x > \bar{x}$). Considering the latter, we have, with $\lambda \geq 0$,

Proof of Cramér's Theorem

To see the large deviations upper bound, we note first that the strict convexity of $I(\cdot)$ implies that the latter is strictly monotone increasing (resp. decreasing) on $[\bar{x}, \infty)$ (resp. $(-\infty, \bar{x}]$)). It is therefore enough to prove the upper bound on intervals of the form $(-\infty, x]$ (with $x < \bar{x}$) or $[x, \infty)$ (with $x > \bar{x}$). Considering the latter, we have, with $\lambda \geq 0$,

$$\mathbb{P}(S_n/n \geq x) \leq \mathbb{E}(e^{-n\lambda x + \lambda S_n}) = e^{-n(\lambda x - \Lambda(\lambda))}$$

Proof of Cramér's Theorem

To see the large deviations upper bound, we note first that the strict convexity of $I(\cdot)$ implies that the latter is strictly monotone increasing (resp. decreasing) on $[\bar{x}, \infty)$ (resp. $(-\infty, \bar{x}]$)). It is therefore enough to prove the upper bound on intervals of the form $(-\infty, x]$ (with $x < \bar{x}$) or $[x, \infty)$ (with $x > \bar{x}$). Considering the latter, we have, with $\lambda \geq 0$,

$$\mathbb{P}(S_n/n \geq x) \leq \mathbb{E}(e^{-n\lambda x + \lambda S_n}) = e^{-n(\lambda x - \Lambda(\lambda))}$$

where the independence of the X_i's was used in the last equality. Choosing $\lambda = \lambda_x$ completes the proof of the large deviations upper bound.

Proof of Cramér's Theorem

To see the lower bound, it is enough to show that

Proof of Cramér's Theorem

To see the lower bound, it is enough to show that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n/n\in(x-\delta,x+\delta))=-I(x)$$

Proof of Cramér's Theorem

To see the lower bound, it is enough to show that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n/n\in(x-\delta,x+\delta))=-I(x)$$

To see this, introduce the probability distribution ν by setting

Proof of Cramér's Theorem

To see the lower bound, it is enough to show that

$$\lim_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n/n \in (x-\delta, x+\delta)) = -I(x)$$

To see this, introduce the probability distribution ν by setting $d\nu/d\mathbb{P}_1(y) = e^{\lambda_x y - \Lambda(\lambda_x)}$. Then,

Proof of Cramér's Theorem

To see the lower bound, it is enough to show that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n/n\in(x-\delta,x+\delta)) = -I(x)$$

To see this, introduce the probability distribution ν by setting $d\nu/d\mathbb{P}_1(y) = e^{\lambda_x y - \Lambda(\lambda_x)}$. Then,

$$\begin{aligned}\mathbb{E}_\nu(X_1) &= \int y e^{\lambda_x y-\Lambda(\lambda_x)}\,d\mathbb{P}_1(y)\\ &= e^{-\Lambda(\lambda_x)}\frac{d}{d\lambda}\,e^{\Lambda(\lambda)}\Big|_{\lambda=\lambda_x}\\ &= \Lambda'(\lambda_x) = x\end{aligned}$$

so, by the definition of λ_x

Proof of Cramér's Theorem

so, by the definition of λ_x

$$\begin{aligned}
\mathbb{P}(S_n/n \in (x-\delta, x+\delta)) &= \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\mathbb{P}_1(x_i) \\
&= \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} e^{-\lambda_x \sum_{i=1}^n x_i + n\Lambda(\lambda_x)} \prod_{i=1}^n d\nu(x_i) \\
&\geq e^{-n(\lambda_x\delta + x\lambda_x - \Lambda(\lambda_x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\nu(x_i) \\
&= e^{-n(\lambda_x\delta + I(x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\nu(x_i)
\end{aligned}$$

Proof of Cramér's Theorem

so, by the definition of λ_x

$$\mathbb{P}(S_n/n \in (x-\delta, x+\delta)) = \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\mathbb{P}_1(x_i)$$

$$= \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} e^{-\lambda_x \sum_{i=1}^n x_i + n\Lambda(\lambda_x)} \prod_{i=1}^n d\nu(x_i)$$

$$\geq e^{-n(\lambda_x\delta + x\lambda_x - \Lambda(\lambda_x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\nu(x_i)$$

$$= e^{-n(\lambda_x\delta + I(x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta,x+\delta)} \prod_{i=1}^n d\nu(x_i)$$

Since $E_\nu(X_1) = x$, the law of large numbers implies that for any $\delta > 0$, the last integral in converges to 1 as $n \to \infty$.

Proof of Cramér's Theorem

so, by the definition of λ_x

$$\begin{aligned}
\mathbb{P}(S_n/n \in (x-\delta, x+\delta)) &= \int_{\sum_{i=1}^n x_i/n \in (x-\delta, x+\delta)} \prod_{i=1}^n d\mathbb{P}_1(x_i) \\
&= \int_{\sum_{i=1}^n x_i/n \in (x-\delta, x+\delta)} e^{-\lambda_x \sum_{i=1}^n x_i + n\Lambda(\lambda_x)} \prod_{i=1}^n d\nu(x_i) \\
&\geq e^{-n(\lambda_x \delta + x\lambda_x - \Lambda(\lambda_x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta, x+\delta)} \prod_{i=1}^n d\nu(x_i) \\
&= e^{-n(\lambda_x \delta + I(x))} \int_{\sum_{i=1}^n x_i/n \in (x-\delta, x+\delta)} \prod_{i=1}^n d\nu(x_i)
\end{aligned}$$

Since $E_\nu(X_1) = x$, the law of large numbers implies that for any $\delta > 0$, the last integral in converges to 1 as $n \to \infty$. Taking the limits $n \to \infty$ first and then $\delta \to 0$ completes the proof of the lower bound.

Some Examples

We'll consider three examples, two where we can compute the rate function directly

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer)

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1.

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1. Then S_n is Gaussian with mean 0 and second moment n^{-1} so

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1. Then S_n is Gaussian with mean 0 and second moment n^{-1} so

$$d\mathbb{P}_n(x) = \sqrt{\frac{2\pi}{n}} e^{-nx^2/2} dx$$

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1. Then S_n is Gaussian with mean 0 and second moment n^{-1} so

$$d\mathbb{P}_n(x) = \sqrt{\frac{2\pi}{n}} e^{-nx^2/2} dx$$

which, by the earlier result on $\mathbb{R}^\nu$ measures has an LDP with speed n and rate $I(x) = \frac{1}{2}x^2$.

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1. Then S_n is Gaussian with mean 0 and second moment n^{-1} so

$$d\mathbb{P}_n(x) = \sqrt{\frac{2\pi}{n}} e^{-nx^2/2} dx$$

which, by the earlier result on $\mathbb{R}^\nu$ measures has an LDP with speed n and rate $I(x) = \frac{1}{2}x^2$. By completing the square, one computes $\Lambda(\lambda) = \frac{1}{2}\lambda^2$

Some Examples

We'll consider three examples, two where we can compute the rate function directly (and check that Cramér gives the same answer) and a last one relevant to the final lecture.

Example 1 (Gaussian) Let ξ be a Gaussian with mean 0 and second moment 1. Then S_n is Gaussian with mean 0 and second moment n^{-1} so

$$d\mathbb{P}_n(x) = \sqrt{\frac{2\pi}{n}} e^{-nx^2/2} dx$$

which, by the earlier result on $\mathbb{R}^\nu$ measures has an LDP with speed n and rate $I(x) = \frac{1}{2}x^2$. By completing the square, one computes $\Lambda(\lambda) = \frac{1}{2}\lambda^2$ so $\lambda x - \Lambda(\lambda) = \frac{1}{2}x^2 - \frac{1}{2}(x-\lambda)^2$ and the sup is exactly $\frac{1}{2}x^2$.

Example 2 (Cauchy distribution) Let ξ have distribution $\frac{1}{\pi}\frac{1}{1+x^2}\,dx$ which has numerous pathological features.

Some Examples

Example 2 (Cauchy distribution) Let ξ have distribution $\frac{1}{\pi}\frac{1}{1+x^2}\,dx$ which has numerous pathological features. One of the more shocking is that the distribution of S_n is independent of n and is just the Cauchy distribution!

Some Examples

Example 2 (Cauchy distribution) Let ξ have distribution $\frac{1}{\pi}\frac{1}{1+x^2}\,dx$ which has numerous pathological features. One of the more shocking is that the distribution of S_n is independent of n and is just the Cauchy distribution! The law of large numbers fails but that doesn't violate the theorems since $\mathbb{E}(|\xi|) = \infty$.

Some Examples

Example 2 (Cauchy distribution) Let ξ have distribution $\frac{1}{\pi}\frac{1}{1+x^2}\,dx$ which has numerous pathological features. One of the more shocking is that the distribution of S_n is independent of n and is just the Cauchy distribution! The law of large numbers fails but that doesn't violate the theorems since $\mathbb{E}(|\xi|) = \infty$. A little thought shows that since every open set has positive probability independently of n, we have that an LDP holds with $I \equiv 0$.

Example 2 (Cauchy distribution) Let ξ have distribution $\frac{1}{\pi}\frac{1}{1+x^2}\,dx$ which has numerous pathological features. One of the more shocking is that the distribution of S_n is independent of n and is just the Cauchy distribution! The law of large numbers fails but that doesn't violate the theorems since $\mathbb{E}(|\xi|) = \infty$. A little thought shows that since every open set has positive probability independently of n, we have that an LDP holds with $I \equiv 0$.

Clearly, $\Lambda(\lambda) = \infty$ for $\lambda \neq 0$ and $\Lambda(0) = 0$. For any x, the $\sup$ of $\lambda x - \Lambda(\lambda)$ occurs at $\lambda = 0$ so the Legendre transform is $\equiv 0$ showing the above calculation is consistent with Cramér's theorem.

Some Examples

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^{\infty}$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

Some Examples

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^{\infty}$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

$$\Lambda(\lambda) = \log\left(\int_0^\infty e^{\lambda x} e^{-x}\, dx\right) = \begin{cases} -\log(1-\lambda), & \text{if } \lambda < 1 \\ \infty, & \text{if } \lambda \geq 1 \end{cases}$$

Some Examples

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^{\infty}$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

$$\Lambda(\lambda) = \log\left(\int_0^\infty e^{\lambda x}e^{-x}\,dx\right) = \begin{cases} -\log(1-\lambda), & \text{if } \lambda < 1 \\ \infty, & \text{if } \lambda \geq 1 \end{cases}$$

For $x \leq 0$, taking $\lambda \to -\infty$ in $\lambda x - \Lambda(\lambda)$, we see that $I(x) = \infty$. If $x > 0$, the λ derivative of $\lambda x - \Lambda(\lambda)$ vanishes at $\lambda = 1 - x^{-1}$ at which point $\lambda x - \Lambda(\lambda)$ has the value $x - 1 - \log(x)$. Thus

Some Examples

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^\infty$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

$$\Lambda(\lambda) = \log\left(\int_0^\infty e^{\lambda x}e^{-x}\,dx\right) = \begin{cases} -\log(1-\lambda), & \text{if } \lambda < 1 \\ \infty, & \text{if } \lambda \geq 1 \end{cases}$$

For $x \leq 0$, taking $\lambda \to -\infty$ in $\lambda x - \Lambda(\lambda)$, we see that $I(x) = \infty$. If $x > 0$, the λ derivative of $\lambda x - \Lambda(\lambda)$ vanishes at $\lambda = 1 - x^{-1}$ at which point $\lambda x - \Lambda(\lambda)$ has the value $x - 1 - \log(x)$. Thus

$$\varphi(x) \equiv I(x) = \begin{cases} x - 1 - \log(x), & \text{if } x > 0 \\ \infty, & \text{if } x \leq 0 \end{cases}$$

Some Examples

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^{\infty}$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

$$\Lambda(\lambda) = \log\left(\int_0^\infty e^{\lambda x}e^{-x}\,dx\right) = \begin{cases} -\log(1-\lambda), & \text{if } \lambda < 1 \\ \infty, & \text{if } \lambda \geq 1 \end{cases}$$

For $x \leq 0$, taking $\lambda \to -\infty$ in $\lambda x - \Lambda(\lambda)$, we see that $I(x) = \infty$. If $x > 0$, the λ derivative of $\lambda x - \Lambda(\lambda)$ vanishes at $\lambda = 1 - x^{-1}$ at which point $\lambda x - \Lambda(\lambda)$ has the value $x - 1 - \log(x)$. Thus

$$\varphi(x) \equiv I(x) = \begin{cases} x - 1 - \log(x), & \text{if } x > 0 \\ \infty, & \text{if } x \leq 0 \end{cases}$$

is the (good) rate function.

Example 3 (Exponential distribution) Relevant to our considerations later is the average of exponential random variables. So let $\{X_j\}_{j=1}^{\infty}$ be independent, identically distributed random variables (iidrv) with density $\chi_{[0,\infty)}(x)e^{-x}dx$. The cumulant generating function is

$$\Lambda(\lambda) = \log\left(\int_0^\infty e^{\lambda x} e^{-x}\,dx\right) = \begin{cases} -\log(1-\lambda), & \text{if } \lambda < 1 \\ \infty, & \text{if } \lambda \geq 1 \end{cases}$$

For $x \leq 0$, taking $\lambda \to -\infty$ in $\lambda x - \Lambda(\lambda)$, we see that $I(x) = \infty$. If $x > 0$, the λ derivative of $\lambda x - \Lambda(\lambda)$ vanishes at $\lambda = 1 - x^{-1}$ at which point $\lambda x - \Lambda(\lambda)$ has the value $x - 1 - \log(x)$. Thus

$$\varphi(x) \equiv I(x) = \begin{cases} x - 1 - \log(x), & \text{if } x > 0 \\ \infty, & \text{if } x \leq 0 \end{cases}$$

is the (good) rate function. It is no coincidence as we'll see that $G(a) = \varphi(a^2)$.

Some Examples

We summarize the combination of this calculation and Cramér's Theorem in the theorem below which we'll need in the last lecture. The gamma distribution is the measure given by

Some Examples

We summarize the combination of this calculation and Cramér's Theorem in the theorem below which we'll need in the last lecture. The gamma distribution is the measure given by

$$dG_{\alpha,\beta}(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)}\, \chi_{[0,\infty)}(x)\, dx$$

Some Examples

We summarize the combination of this calculation and Cramér's Theorem in the theorem below which we'll need in the last lecture. The gamma distribution is the measure given by

$$dG_{\alpha,\beta}(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-x\beta}}{\Gamma(\alpha)}\, \chi_{[0,\infty)}(x)\, dx$$

For exponential iidrv, $n^{-1}\sum_{j=1}^n X_n$ has distribution $G_{n-1,n}$, so this example allows one to also read off a LDP for suitable gamma distributions.

Some Examples

Theorem Let ℓ_N be integers with $\lim_{N\to\infty} N^{-1}\ell_N = \alpha > 0$.

Some Examples

Theorem Let ℓ_N be integers with $\lim_{N\to\infty} N^{-1}\ell_N = \alpha > 0$. Then $Y_N \equiv N^{-1}\sum_{j=1}^{\ell_N} X_j$ with X_j iid exponential random variables obeys a LDP with speed N and rate function

Some Examples

Theorem Let ℓ_N be integers with $\lim_{N\to\infty} N^{-1}\ell_N = \alpha > 0$. Then $Y_N \equiv N^{-1}\sum_{j=1}^{\ell_N} X_j$ with X_j iid exponential random variables obeys a LDP with speed N and rate function

$$\varphi_\alpha(y) \equiv \alpha\varphi(y/\alpha) = y - \alpha - \alpha\log(y/\alpha)$$

Some Examples

Theorem Let ℓ_N be integers with $\lim_{N\to\infty} N^{-1}\ell_N = \alpha > 0$. Then $Y_N \equiv N^{-1}\sum_{j=1}^{\ell_N} X_j$ with X_j iid exponential random variables obeys a LDP with speed N and rate function

$$\varphi_\alpha(y) \equiv \alpha\varphi(y/\alpha) = y - \alpha - \alpha\log(y/\alpha)$$

Remark This goes beyond the direct use of Cramér in two ways. First, we note that if real valued Z_N have a LDP with speed N and rate I, then αZ_N has a LDP with speed αN and rate $\alpha I(\cdot/\alpha)$ by a trivial calculation. Secondly, if $\alpha_N = \ell_N/N$ and $\alpha_N \to 1$, then $\alpha_N^{-1} Y_N$ has a LDP with speed N and rate I if Y_N does and the rate function is continuous.

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps.

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over,

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over, but in a few places we'll need the following:

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over, but in a few places we'll need the following:

Contraction Principle Let X and Y be Polish spaces and $f : X \to Y$ a continuous function onto Y.

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over, but in a few places we'll need the following:

Contraction Principle Let X and Y be Polish spaces and $f : X \to Y$ a continuous function onto Y. Suppose $\{\mathbb{P}_N\}_{N=1}^{\infty}$ is a family of probability measures on X that obeys a LDP with speed a_N and good rate function I. Define on Y the function

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over, but in a few places we'll need the following:

Contraction Principle Let X and Y be Polish spaces and $f : X \to Y$ a continuous function onto Y. Suppose $\{\mathbb{P}_N\}_{N=1}^{\infty}$ is a family of probability measures on X that obeys a LDP with speed a_N and good rate function I. Define on Y the function

$$I^{(f)}(y) = \inf\{I(x) \mid f(x) = y\}$$

Statement of Contraction Principle

Next, we discuss a result known as the contraction principle which allows one to pull a LDP over under continuous maps. For most of our basic situation, the maps are homeomorphisms so it is trivial that LDP's carry over, but in a few places we'll need the following:

Contraction Principle Let X and Y be Polish spaces and $f : X \to Y$ a continuous function onto Y. Suppose $\{\mathbb{P}_N\}_{N=1}^{\infty}$ is a family of probability measures on X that obeys a LDP with speed a_N and good rate function I. Define on Y the function

$$I^{(f)}(y) = \inf\{I(x) \mid f(x) = y\}$$

Then the family of measures on Y defined by $\mathbb{P}_N^{(f)}(A) = \mathbb{P}_N(f^{-1}[A])$ obeys a LDP with speed a_N and good rate function $I^{(f)}$.

Proof of Contraction Principle

A simple argument shows that $I^{(f)}$ is a good rate function.
If A is open (resp. closed), so is $f^{-1}[A]$ and

Proof of Contraction Principle

A simple argument shows that $I^{(f)}$ is a good rate function.
If A is open (resp. closed), so is $f^{-1}[A]$ and

$$\inf_{y\in A} I^{(f)}(y) = \inf_{x\in f^{-1}[A]} I(x)$$

Proof of Contraction Principle

A simple argument shows that $I^{(f)}$ is a good rate function. If A is open (resp. closed), so is $f^{-1}[A]$ and

$$\inf_{y \in A} I^{(f)}(y) = \inf_{x \in f^{-1}[A]} I(x)$$

so the LDP bounds for $\mathbb{P}_N$ carry over to such bounds for $\mathbb{P}_N^{(f)}$.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^{\infty}$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^\infty$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto so that $\pi_{j+1,j}\pi_{j+1} = \pi_j$.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^{\infty}$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto so that $\pi_{j+1,j}\pi_{j+1} = \pi_j$. We require that if $\pi_j(x) = \pi_j(y)$ for $x, y \in X$ and all j, then $x = y$.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^\infty$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto so that $\pi_{j+1,j}\pi_{j+1} = \pi_j$. We require that if $\pi_j(x) = \pi_j(y)$ for $x, y \in X$ and all j, then $x = y$. In the abstract discussion, one just needs to be given X_j and $\pi_{j+1,j}$ and can form X as the subset of $\prod_{j=1}^\infty X_j$ of those $x = (x_j)$ with $\pi_{j+1,j}(x_{j+1}) = x_j$. One puts the product topology on X.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^\infty$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto so that $\pi_{j+1,j}\pi_{j+1} = \pi_j$. We require that if $\pi_j(x) = \pi_j(y)$ for $x, y \in X$ and all j, then $x = y$. In the abstract discussion, one just needs to be given X_j and $\pi_{j+1,j}$ and can form X as the subset of $\prod_{j=1}^\infty X_j$ of those $x = (x_j)$ with $\pi_{j+1,j}(x_{j+1}) = x_j$. One puts the product topology on X. In the cases of interest, X will be explicitly given but it agrees with this abstract construction.

Projective Limit Theorem

The last topic subject in the general LD theory that we want to consider the theory of projective limits of LDP's or at least a very special case – projective limits are indexed by directed sets; we'll only consider the case where the directed set is $\mathbb{Z}_+$. We have Polish spaces $\{X_j\}_{j=1}^{\infty}$ and X and continuous maps $\pi_j : X \to X_j$ and $\pi_{j+1,j} : X_{j+1} \to X_j$ all onto so that $\pi_{j+1,j}\pi_{j+1} = \pi_j$. We require that if $\pi_j(x) = \pi_j(y)$ for $x, y \in X$ and all j, then $x = y$. In the abstract discussion, one just needs to be given X_j and $\pi_{j+1,j}$ and can form X as the subset of $\prod_{j=1}^{\infty} X_j$ of those $x = (x_j)$ with $\pi_{j+1,j}(x_{j+1}) = x_j$. One puts the product topology on X. In the cases of interest, X will be explicitly given but it agrees with this abstract construction. For a measure $\mathbb{P}$ on X, define $\pi_j^*(\mathbb{P})$, a measure on X_j by $\pi_j^*(\mathbb{P})(A) = \mathbb{P}(\pi_j^{-1}[A])$.

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner – there is a whole chapter in DZ on the subject.

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner – there is a whole chapter in DZ on the subject.

Projective Limit Theorem Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a family of measures on X.

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner – there is a whole chapter in DZ on the subject.

Projective Limit Theorem Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a family of measures on X. Suppose that for each j, $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP with speed a_N and good rate function, I_j on X_j. Let

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner – there is a whole chapter in DZ on the subject.

Projective Limit Theorem Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a family of measures on X. Suppose that for each j, $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP with speed a_N and good rate function, I_j on X_j. Let

$$I(x) = \sup_j \{I_j(\pi_j(x))\}$$

Projective Limit Theorem

Here is a basic theorem due to Dawson–Gärtner – there is a whole chapter in DZ on the subject.

Projective Limit Theorem Let $\{\mathbb{P}_N\}_{N=1}^{\infty}$ be a family of measures on X. Suppose that for each j, $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP with speed a_N and good rate function, I_j on X_j. Let

$$I(x) = \sup_j \{I_j(\pi_j(x))\}$$

Then I is a good rate function and $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP with speed a_N and rate function I.

Projective Limit Theorem

Remarks 1. The converse, i.e. if $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP then so does each $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$, is trivial by the contraction principle.

Projective Limit Theorem

Remarks 1. The converse, i.e. if $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP then so does each $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$, is trivial by the contraction principle.

2. The same idea shows that if $\{\pi_{j+1}^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP, so does $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ and

Projective Limit Theorem

Remarks 1. The converse, i.e. if $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP then so does each $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$, is trivial by the contraction principle.

2. The same idea shows that if $\{\pi_{j+1}^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP, so does $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ and

$$I_j(x) = \inf\{I_{j+1}(y) \mid \pi_j(y) = x\}$$

Remarks 1. The converse, i.e. if $\{\mathbb{P}_N\}_{N=1}^{\infty}$ obeys a LDP then so does each $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$, is trivial by the contraction principle.

2. The same idea shows that if $\{\pi_{j+1}^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ obeys a LDP, so does $\{\pi_j^*(\mathbb{P}_N)\}_{N=1}^{\infty}$ and

$$I_j(x) = \inf\{I_{j+1}(y) \,|\, \pi_j(y) = x\}$$

which shows that $I_j(\pi_j(x))$ is monotone in j so the $\sup$ in the formula for I is a limit.

Some Examples

Example $[\mathbb{R}^\infty]$ Take $X_j = \mathbb{R}^j$,
$X = \mathbb{R}^\infty = \{(x_1, x_2, \dots) \mid x_j \in \mathbb{R}\}$ which is a Polish space and $\pi_j(x)_k = x_k$ for $k = 1, \dots, j$.

Some Examples

Example [$\mathbb{R}^\infty$] Take $X_j = \mathbb{R}^j$, $X = \mathbb{R}^\infty = \{(x_1, x_2, \dots) \mid x_j \in \mathbb{R}\}$ which is a Polish space and $\pi_j(x)_k = x_k$ for $k = 1, \dots, j$. The projective limit theorem says that to prove a LDP for X, we need only prove it for the finite dimensional $\mathbb{R}^j$.

Some Examples

Example $[\mathbb{R}^\infty]$ Take $X_j = \mathbb{R}^j$, $X = \mathbb{R}^\infty = \{(x_1, x_2, \dots) \mid x_j \in \mathbb{R}\}$ which is a Polish space and $\pi_j(x)_k = x_k$ for $k = 1, \dots, j$. The projective limit theorem says that to prove a LDP for X, we need only prove it for the finite dimensional $\mathbb{R}^j$.

Example $[\mathcal{M}_{+,1}(\partial\mathbb{D})]$ Let $\mathbb{P}$ be a measure on $\mathcal{M}_{+,1}(\partial\mathbb{D})$, the probability measures on the unit circle.

Some Examples

Example [$\mathbb{R}^\infty$] Take $X_j = \mathbb{R}^j$, $X = \mathbb{R}^\infty = \{(x_1, x_2, \dots) \mid x_j \in \mathbb{R}\}$ which is a Polish space and $\pi_j(x)_k = x_k$ for $k = 1, \dots, j$. The projective limit theorem says that to prove a LDP for X, we need only prove it for the finite dimensional $\mathbb{R}^j$.

Example [$\mathcal{M}_{+,1}(\partial\mathbb{D})$] Let $\mathbb{P}$ be a measure on $\mathcal{M}_{+,1}(\partial\mathbb{D})$, the probability measures on the unit circle. Given $\mu \in \mathcal{M}_{+,1}(\partial\mathbb{D})$ and $j = 1, 2, \dots$, let $\pi_j(\mu)$ be the point in $\mathbb{R}^{2^j}$ with coordinates $\mu(I_k^{(j)}), k = 1, \dots, 2^j$ where $I_k^{(j)} = \{e^{2\pi i\theta} \mid \frac{k-1}{2^j} \leq \theta < \frac{k}{2^j}\}$.

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace.

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace. The $\pi_j(\mu)$ determine $\mu(\{e^{2\pi i\theta} \,|\, 0 \le \theta < \frac{k}{2^j}\})$ and so μ.

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace. The $\pi_j(\mu)$ determine $\mu(\{e^{2\pi i\theta} \mid 0 \le \theta < \frac{k}{2^j}\})$ and so μ. Clearly as $j \to \infty$, $\pi_j(\mu)$ converges weakly to μ.

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace. The $\pi_j(\mu)$ determine $\mu(\{e^{2\pi i\theta} \,|\, 0 \le \theta < \frac{k}{2^j}\})$ and so μ. Clearly as $j \to \infty$, $\pi_j(\mu)$ converges weakly to μ.

In this case

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace. The $\pi_j(\mu)$ determine $\mu(\{e^{2\pi i\theta} \mid 0 \le \theta < \frac{k}{2^j}\})$ and so μ. Clearly as $j \to \infty$, $\pi_j(\mu)$ converges weakly to μ.

In this case

$$\pi_{j+1,j}(y)_\ell = y_{2\ell-1} + y_\ell \qquad \ell = 1, \dots, 2^j$$

Some Examples

Realizing $\mathbb{R}^{2^j}$ as a set of measures, we can think of

$$\pi_j(\mu) = \sum_{k=1}^{2^j} \mu(I_k^{(j)}) 2^j \chi_{I_k^{(j)}}(x)\, dx$$

Thus $\mathbb{P}$ induces a measure $\pi_j^*(\mathbb{P})$ on either $\mathbb{R}^{2^j}$ or on $\mathcal{M}_{+,1}(\partial\mathbb{D})$ supported on a 2^j–dimensional subspace. The $\pi_j(\mu)$ determine $\mu(\{e^{2\pi i\theta} \mid 0 \leq \theta < \frac{k}{2^j}\})$ and so μ. Clearly as $j \to \infty$, $\pi_j(\mu)$ converges weakly to μ.

In this case

$$\pi_{j+1,j}(y)_\ell = y_{2\ell-1} + y_\ell \qquad \ell = 1, \dots, 2^j$$

Thus, to get a LDP for $\mathcal{M}_{+,1}(\partial\mathbb{D})$, we need only prove 2^j–dimensional LDPs.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work).

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic,

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic, so for a.e. $U \in \mathbb{U}(n)$, it defines a spectral measure on $\partial\mathbb{C}$, an n–point measure,

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic, so for a.e. $U \in \mathbb{U}(n)$, it defines a spectral measure on $\partial\mathbb{C}$, an n–point measure, and so n Verblunsky coefficients.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic, so for a.e. $U \in \mathbb{U}(n)$, it defines a spectral measure on $\partial\mathbb{C}$, an n–point measure, and so n Verblunsky coefficients.

Haar measure thus induces a measure on the $2n - 1$ (real) dimensional set of possible Verblunsky coefficients.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic, so for a.e. $U \in \mathbb{U}(n)$, it defines a spectral measure on $\partial\mathbb{C}$, an n–point measure, and so n Verblunsky coefficients.

Haar measure thus induces a measure on the $2n-1$ (real) dimensional set of possible Verblunsky coefficients. Killip and Nenciu asked and answered what this probability measure is.

Statement of the Theorem

I want to get a head start on the next lecture by ending this with at least part of the discussion of a result we'll need next time. Let $\mathbb{U}(n)$ be the $n \times n$ unitary matrices. There is a unique probability measure on $\mathbb{U}(n)$ invariant under both left and right multiplication by any unitary. It is called Haar measure (although for Lie groups, Hurwitz found the invariant measure long before Haar's work). For any fixed unit vector, $\varphi \in \mathbb{C}^n$, it easy to see that φ is cyclic, so for a.e. $U \in \mathbb{U}(n)$, it defines a spectral measure on $\partial\mathbb{C}$, an n–point measure, and so n Verblunsky coefficients.

Haar measure thus induces a measure on the $2n-1$ (real) dimensional set of possible Verblunsky coefficients. Killip and Nenciu asked and answered what this probability measure is. They were motivated by a paper of Dmitriu and Edelman who had asked and answered the analogous question for GUE and Jacobi parameters.

Statement of the Theorem

Killip–Nenciu Theorem The measure induced by Haar measure on $\mathbb{U}(n)$ on the Verblunsky coefficients $\alpha_0, \dots, \alpha_{n-2} \in \mathbb{D}, \alpha_{n-1} \in \partial\mathbb{D}$ is given given by

Statement of the Theorem

Killip–Nenciu Theorem The measure induced by Haar measure on $\mathbb{U}(n)$ on the Verblunsky coefficients $\alpha_0, \ldots, \alpha_{n-2} \in \mathbb{D}, \alpha_{n-1} \in \partial\mathbb{D}$ is given given by

$$d\widetilde{\mathbb{P}}_n(\alpha_0, \ldots, \alpha_{n-1}) = \prod_{j=0}^{n-1} d\kappa_{n-2-j}(\alpha_j)$$

$$d\kappa_\ell(\alpha) = \frac{\ell+1}{\pi}(1-|\alpha|^2)^\ell d^2\alpha \quad \text{on } \mathbb{D};\ \ell \geq 0$$

$$d\kappa_{-1}(\alpha = e^{i\theta}) = \frac{d\theta}{2\pi} \quad \text{on } \partial\mathbb{D}$$

Statement of the Theorem

Killip–Nenciu Theorem The measure induced by Haar measure on $\mathbb{U}(n)$ on the Verblunsky coefficients $\alpha_0, \dots, \alpha_{n-2} \in \mathbb{D}, \alpha_{n-1} \in \partial\mathbb{D}$ is given given by

$$d\widetilde{\mathbb{P}}_n(\alpha_0, \dots, \alpha_{n-1}) = \prod_{j=0}^{n-1} d\kappa_{n-2-j}(\alpha_j)$$

$$d\kappa_\ell(\alpha) = \frac{\ell+1}{\pi}(1-|\alpha|^2)^\ell d^2\alpha \quad \text{on } \mathbb{D};\ \ell \geq 0$$

$$d\kappa_{-1}(\alpha = e^{i\theta}) = \frac{d\theta}{2\pi} \quad \text{on } \partial\mathbb{D}$$

This says the α_j are independent.

Statement of the Theorem

Killip–Nenciu Theorem The measure induced by Haar measure on $\mathbb{U}(n)$ on the Verblunsky coefficients $\alpha_0, \ldots, \alpha_{n-2} \in \mathbb{D}, \alpha_{n-1} \in \partial\mathbb{D}$ is given given by

$$d\widetilde{\mathbb{P}}_n(\alpha_0, \ldots, \alpha_{n-1}) = \prod_{j=0}^{n-1} d\kappa_{n-2-j}(\alpha_j)$$

$$d\kappa_\ell(\alpha) = \frac{\ell+1}{\pi}(1-|\alpha|^2)^\ell d^2\alpha \quad \text{on } \mathbb{D};\ \ell \geq 0$$

$$d\kappa_{-1}(\alpha = e^{i\theta}) = \frac{d\theta}{2\pi} \quad \text{on } \partial\mathbb{D}$$

This says the α_j are independent. The distribution $d\kappa_\ell$ is exactly the distribution of z_1 if $\mathbf{z} \equiv (z_1, \ldots, z_{\ell+2}) \in \mathbb{C}^{\ell+2}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{\ell+2}$.

Statement of the Theorem

Killip–Nenciu Theorem The measure induced by Haar measure on $\mathbb{U}(n)$ on the Verblunsky coefficients $\alpha_0, \ldots, \alpha_{n-2} \in \mathbb{D}, \alpha_{n-1} \in \partial\mathbb{D}$ is given given by

$$d\widetilde{\mathbb{P}}_n(\alpha_0, \ldots, \alpha_{n-1}) = \prod_{j=0}^{n-1} d\kappa_{n-2-j}(\alpha_j)$$

$$d\kappa_\ell(\alpha) = \frac{\ell+1}{\pi}(1-|\alpha|^2)^\ell d^2\alpha \quad \text{on } \mathbb{D};\ \ell \geq 0$$

$$d\kappa_{-1}(\alpha = e^{i\theta}) = \frac{d\theta}{2\pi} \quad \text{on } \partial\mathbb{D}$$

This says the α_j are independent. The distribution $d\kappa_\ell$ is exactly the distribution of z_1 if $\mathbf{z} \equiv (z_1, \ldots, z_{\ell+2}) \in \mathbb{C}^{\ell+2}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{\ell+2}$. That is, α_0 is distributed like the first component of the unit sphere in $\mathbb{C}^n$, α_1 in $\mathbb{C}^{n-1}$, etc.

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof.

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy.

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy. The subtle step will be the independence so we begin with the tools needed to see that.

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy. The subtle step will be the independence so we begin with the tools needed to see that.

We start with a natural matrix representation of the matrix with given Verblunsky coefficients.

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy. The subtle step will be the independence so we begin with the tools needed to see that.

We start with a natural matrix representation of the matrix with given Verblunsky coefficients. Given a n–point probability measure μ on $\partial\mathbb{D}$, we define the *GGT matrix*, $\{\mathcal{G}_{k\ell}^{(n)}(d\mu)\}_{0\leq k,\ell<\infty}$ by

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy. The subtle step will be the independence so we begin with the tools needed to see that.

We start with a natural matrix representation of the matrix with given Verblunsky coefficients. Given a n–point probability measure μ on $\partial\mathbb{D}$, we define the *GGT matrix*, $\{\mathcal{G}_{k\ell}^{(n)}(d\mu)\}_{0\le k,\ell<\infty}$ by

$$\mathcal{G}_{k\ell}^{(n)}(d\mu) = \langle \varphi_k, z\varphi_\ell \rangle \qquad 0 \le k,\ell < \infty$$

GGT Matrix

I want to sketch a proof of this result due to Breuer, Simon and Zeitouni that is a variant of the original KN proof. As we'll see, that α_1 is distributed as the first component of the unit sphere in $\mathbb{C}^n$ will be easy. The subtle step will be the independence so we begin with the tools needed to see that.

We start with a natural matrix representation of the matrix with given Verblunsky coefficients. Given a n–point probability measure μ on $\partial\mathbb{D}$, we define the *GGT matrix*, $\{\mathcal{G}^{(n)}_{k\ell}(d\mu)\}_{0\le k,\ell<\infty}$ by

$$\mathcal{G}^{(n)}_{k\ell}(d\mu) = \langle \varphi_k, z\varphi_\ell\rangle \qquad 0 \le k,\ell < \infty$$

GGT is a name I introduced in OPUC1 after Geronimus (who had it first in his work on OPUC), Gragg (who rediscovered it in work in numerical linear algebra) and Teplyaev (who used it in his study of Anderson localization for OPUC).

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

$$\mathcal{G}_{k\ell}^{(n)}(\{\alpha_j\}_{j=0}^{n-1}) = \begin{cases} -\bar{\alpha}_\ell \alpha_{k-1} \prod_{j=k}^{\ell-1} \rho_j & 0 \le k \le \ell \\ \rho_\ell & k = \ell + 1 \\ 0 & k \ge \ell + 2 \end{cases}$$

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

$$\mathcal{G}_{k\ell}^{(n)}(\{\alpha_j\}_{j=0}^{n-1}) = \begin{cases} -\bar{\alpha}_\ell \alpha_{k-1} \prod_{j=k}^{\ell-1} \rho_j & 0 \leq k \leq \ell \\ \rho_\ell & k = \ell + 1 \\ 0 & k \geq \ell + 2 \end{cases}$$

where, as usual, $\rho_\ell = (1 - |\alpha_\ell|^2)^{1/2}$.

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

$$\mathcal{G}_{k\ell}^{(n)}(\{\alpha_j\}_{j=0}^{n-1}) = \begin{cases} -\bar{\alpha}_\ell \alpha_{k-1} \prod_{j=k}^{\ell-1} \rho_j & 0 \le k \le \ell \\ \rho_\ell & k = \ell + 1 \\ 0 & k \ge \ell + 2 \end{cases}$$

where, as usual, $\rho_\ell = (1 - |\alpha_\ell|^2)^{1/2}$. The last which says $\mathcal{G}$ is a Hessenberg matrix follows from the obvious fact that $z\varphi_\ell$ is a polynomial of degree $\ell + 1$.

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

$$\mathcal{G}_{k\ell}^{(n)}(\{\alpha_j\}_{j=0}^{n-1}) = \begin{cases} -\bar{\alpha}_\ell \alpha_{k-1} \prod_{j=k}^{\ell-1} \rho_j & 0 \le k \le \ell \\ \rho_\ell & k = \ell + 1 \\ 0 & k \ge \ell + 2 \end{cases}$$

where, as usual, $\rho_\ell = (1 - |\alpha_\ell|^2)^{1/2}$. The last which says $\mathcal{G}$ is a Hessenberg matrix follows from the obvious fact that $z\varphi_\ell$ is a polynomial of degree $\ell + 1$. Thus $\mathcal{G}(\{\alpha_j\}_{j=0}^{n})$ is given by

GGT Matrix

If $\{\alpha_j\}_{j=0}^{n-1}$ are the Verblunsky coefficients of μ, set $\alpha_{-1} = -1$ and find that:

$$\mathcal{G}_{k\ell}^{(n)}(\{\alpha_j\}_{j=0}^{n-1}) = \begin{cases} -\bar{\alpha}_\ell \alpha_{k-1} \prod_{j=k}^{\ell-1} \rho_j & 0 \le k \le \ell \\ \rho_\ell & k = \ell + 1 \\ 0 & k \ge \ell + 2 \end{cases}$$

where, as usual, $\rho_\ell = (1 - |\alpha_\ell|^2)^{1/2}$. The last which says $\mathcal{G}$ is a Hessenberg matrix follows from the obvious fact that $z\varphi_\ell$ is a polynomial of degree $\ell + 1$. Thus $\mathcal{G}(\{\alpha_j\}_{j=0}^{n})$ is given by

$$\begin{pmatrix} \bar{\alpha}_0 & \bar{\alpha}_1\rho_0 & \bar{\alpha}_2\rho_0\rho_1 & \bar{\alpha}_3\rho_0\rho_1\rho_2 & \dots \\ \rho_0 & -\bar{\alpha}_1\alpha_0 & -\bar{\alpha}_2\alpha_0\rho_1 & -\bar{\alpha}_3\alpha_0\rho_1\rho_2 & \dots \\ 0 & \rho_1 & -\bar{\alpha}_2\alpha_1 & -\bar{\alpha}_3\alpha_1\rho_2 & \dots \\ 0 & 0 & \rho_2 & -\bar{\alpha}_3\alpha_2 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*).

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \ldots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \ldots, z\varphi_{n-1})$.

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \ldots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \ldots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \ldots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \ldots, \varphi_n)$$

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$).

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

$$z\varphi_j = \rho_j \varphi_{j+1} + \bar{\alpha}_j \varphi_j^*$$

$$\varphi_{j+1}^* = -\alpha_j \varphi_{j+1} + \rho_j \varphi_j^*$$

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \ldots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \ldots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \ldots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \ldots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

$$z\varphi_j = \rho_j \varphi_{j+1} + \bar{\alpha}_j \varphi_j^*$$

$$\varphi_{j+1}^* = -\alpha_j \varphi_{j+1} + \rho_j \varphi_j^*$$

Since $\varphi_{j+1} \perp \varphi_j^*$, applying * on degree $j+1$ polynomials,

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \ldots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \ldots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \ldots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \ldots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

$$z\varphi_j = \rho_j \varphi_{j+1} + \bar{\alpha}_j \varphi_j^*$$

$$\varphi_{j+1}^* = -\alpha_j \varphi_{j+1} + \rho_j \varphi_j^*$$

Since $\varphi_{j+1} \perp \varphi_j^*$, applying * on degree $j+1$ polynomials, we see that $\varphi_{j+1}^* \perp z\varphi_j$

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

$$z\varphi_j = \rho_j \varphi_{j+1} + \bar{\alpha}_j \varphi_j^*$$

$$\varphi_{j+1}^* = -\alpha_j \varphi_{j+1} + \rho_j \varphi_j^*$$

Since $\varphi_{j+1} \perp \varphi_j^*$, applying * on degree $j+1$ polynomials, we see that $\varphi_{j+1}^* \perp z\varphi_j$ so the above represents a unitary change of basis on the span of φ_{j+1} and φ_j^*.

AGR Factorization

We turn to my proof of a factorization of $\mathcal{G}$ due to Ammar, Gragg and Reichel (hence *AGR Factorization*). $\mathcal{G}$ can be thought of as the overlap matrix between the two bases: $\mathcal{B}^{(0)} = (\varphi_0, \dots, \varphi_{n-1})$ and $\mathcal{B}^{(n)} = (z\varphi_0, \dots, z\varphi_{n-1})$. Define intermediate bases

$$\mathcal{B}^{(j)} = (z\varphi_0, \dots, z\varphi_{j-1}, \varphi_j^*, \varphi_{j+1}, \dots, \varphi_n)$$

They interpolate since they give us the end points at $j = n$ and $j = 0$ (if one notes that $\varphi_0^* = \varphi_0$). To see they are orthonormal, we first note that

$$z\varphi_j = \rho_j \varphi_{j+1} + \bar{\alpha}_j \varphi_j^*$$

$$\varphi_{j+1}^* = -\alpha_j \varphi_{j+1} + \rho_j \varphi_j^*$$

Since $\varphi_{j+1} \perp \varphi_j^*$, applying * on degree $j+1$ polynomials, we see that $\varphi_{j+1}^* \perp z\varphi_j$ so the above represents a unitary change of basis on the span of φ_{j+1} and φ_j^*. Inductively that implies the $\mathcal{B}^{(j)}$ are orthonormal bases.

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$

for $j \leq n-2$

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$ for $j \leq n-2$ and $\widetilde{\Theta}_{n-1} = \mathbf{1}_{n-1} \oplus \bar{\alpha}_{n-1}$.

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$ for $j \leq n-2$ and $\widetilde{\Theta}_{n-1} = \mathbf{1}_{n-1} \oplus \bar{\alpha}_{n-1}$. Then $\widetilde{\Theta}_j$ is the unitary change of basis from $\mathcal{B}^{(j)}$ to $\mathcal{B}^{(j+1)}$. This implies the AGR factorization:

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$ for $j \leq n-2$ and $\widetilde{\Theta}_{n-1} = \mathbf{1}_{n-1} \oplus \bar{\alpha}_{n-1}$. Then $\widetilde{\Theta}_j$ is the unitary change of basis from $\mathcal{B}^{(j)}$ to $\mathcal{B}^{(j+1)}$. This implies the AGR factorization:

$$\mathcal{G}^{(n)}(\{\alpha\}_{j=0}^{n-1}) = \widetilde{\Theta}_0 \widetilde{\Theta}_1 \dots \widetilde{\Theta}_{n-2} \widetilde{\Theta}_{n-1}$$

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$ for $j \leq n-2$ and $\widetilde{\Theta}_{n-1} = \mathbf{1}_{n-1} \oplus \bar{\alpha}_{n-1}$. Then $\widetilde{\Theta}_j$ is the unitary change of basis from $\mathcal{B}^{(j)}$ to $\mathcal{B}^{(j+1)}$. This implies the AGR factorization:

$$\mathcal{G}^{(n)}(\{\alpha\}_{j=0}^{n-1}) = \widetilde{\Theta}_0 \widetilde{\Theta}_1 \dots \widetilde{\Theta}_{n-2} \widetilde{\Theta}_{n-1}$$

This, in turn, implies the crucial:

AGR Factorization

Let $\Theta(\alpha) = \begin{pmatrix} \bar{\alpha} & \rho \\ \rho & -\alpha \end{pmatrix}$ and $\widetilde{\Theta}_j = \mathbf{1}_j \oplus \Theta(\alpha_j) \oplus \mathbf{1}_{n-j-2}$ for $j \leq n-2$ and $\widetilde{\Theta}_{n-1} = \mathbf{1}_{n-1} \oplus \bar{\alpha}_{n-1}$. Then $\widetilde{\Theta}_j$ is the unitary change of basis from $\mathcal{B}^{(j)}$ to $\mathcal{B}^{(j+1)}$. This implies the AGR factorization:

$$\mathcal{G}^{(n)}(\{\alpha\}_{j=0}^{n-1}) = \widetilde{\Theta}_0 \widetilde{\Theta}_1 \dots \widetilde{\Theta}_{n-2} \widetilde{\Theta}_{n-1}$$

This, in turn, implies the crucial:

$$\mathcal{G}^{(n)}(\{\alpha\}_{j=0}^{n-1}) = (\Theta(\alpha_0) \oplus \mathbf{1}_{n-2}) \left(\mathbf{1}_1 \oplus \mathcal{G}^{(n-1)}(\{\alpha\}_{j=1}^{n-1}) \right)$$

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G.

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection.

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection. Normalized Haar measure, μ_G, induces a natural probability measure, $\mu_{G/H}$, on G/H via

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection. Normalized Haar measure, μ_G, induces a natural probability measure, $\mu_{G/H}$, on G/H via

$$\mu_{G/H}(A) = \mu_G(\pi^{-1}[A])$$

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection. Normalized Haar measure, μ_G, induces a natural probability measure, $\mu_{G/H}$, on G/H via

$$\mu_{G/H}(A) = \mu_G(\pi^{-1}[A])$$

and this measure is clearly invariant under the action of G on G/H.

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection. Normalized Haar measure, μ_G, induces a natural probability measure, $\mu_{G/H}$, on G/H via

$$\mu_{G/H}(A) = \mu_G(\pi^{-1}[A])$$

and this measure is clearly invariant under the action of G on G/H.

Let $\sigma : G/H \to G$ be a choice of representative from each coset, i.e. $\pi(\sigma(x)) = x$. Then $\Sigma : G/H \times H \to G$, defined by $\Sigma(x, h) = \sigma(x)h$, is a bijection.

Diaconis-Shahshahani Formula

Now, let G be a compact group and H a closed subgroup of G. Let $\pi : G \to G/H$ be the canonical projection. Normalized Haar measure, μ_G, induces a natural probability measure, $\mu_{G/H}$, on G/H via

$$\mu_{G/H}(A) = \mu_G(\pi^{-1}[A])$$

and this measure is clearly invariant under the action of G on G/H.

Let $\sigma : G/H \to G$ be a choice of representative from each coset, i.e. $\pi(\sigma(x)) = x$. Then $\Sigma : G/H \times H \to G$, defined by $\Sigma(x, h) = \sigma(x)h$, is a bijection. If one can choose σ to be continuous, then G will be homeomorphic to $G/H \times H$ under Σ and often such a homeomorphism doesn't exist, e.g. if $G = \mathbb{U}(n)$ and $H = \mathbb{U}(n-1)$, so we should avoid the assumption that σ is continuous.

Diaconis-Shahshahani Formula

It is probably true that in general one can make a measurable choice. Since we'll find an explicit such choice below for the case of interest we shall simply suppose that σ is measurable.

Diaconis-Shahshahani Formula

It is probably true that in general one can make a measurable choice. Since we'll find an explicit such choice below for the case of interest we shall simply suppose that σ is measurable.

Diaconis-Shahshahani Formula Suppose σ is measurable. Then under the bijection Σ of $G/H \times H$ and G, the measure $\mu_{G/H} \otimes \mu_H$ goes to μ_G.

Diaconis-Shahshahani Formula

It is probably true that in general one can make a measurable choice. Since we'll find an explicit such choice below for the case of interest we shall simply suppose that σ is measurable.

Diaconis-Shahshahani Formula Suppose σ is measurable. Then under the bijection Σ of $G/H \times H$ and G, the measure $\mu_{G/H} \otimes \mu_H$ goes to μ_G.

To see this, let $U \in G$, $x \in G/H$. Then $\pi(U\sigma(x)) = Ux$ so for some $W_{U,x} \in H$, we have that

Diaconis-Shahshahani Formula

It is probably true that in general one can make a measurable choice. Since we'll find an explicit such choice below for the case of interest we shall simply suppose that σ is measurable.

Diaconis-Shahshahani Formula Suppose σ is measurable. Then under the bijection Σ of $G/H \times H$ and G, the measure $\mu_{G/H} \otimes \mu_H$ goes to μ_G.

To see this, let $U \in G$, $x \in G/H$. Then $\pi(U\sigma(x)) = Ux$ so for some $W_{U,x} \in H$, we have that

$$U\sigma(x) = \sigma(Ux)W_{U,x}$$

Diaconis-Shahshahani Formula

It is probably true that in general one can make a measurable choice. Since we'll find an explicit such choice below for the case of interest we shall simply suppose that σ is measurable.

Diaconis-Shahshahani Formula Suppose σ is measurable. Then under the bijection Σ of $G/H \times H$ and G, the measure $\mu_{G/H} \otimes \mu_H$ goes to μ_G.

To see this, let $U \in G$, $x \in G/H$. Then $\pi(U\sigma(x)) = Ux$ so for some $W_{U,x} \in H$, we have that

$$U\sigma(x) = \sigma(Ux)W_{U,x}$$

so $U\Sigma(x, W) = \Sigma(Ux, W_{U,x}W)$ which, given the invariance of $\mu_{G/H}$ under the action of G and of μ_H under left multiplication by elements of H, implies the image of the product measure is invariant under multiplication by U (by integrating first over W and then x).

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}_1^n$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \ldots, 0)$, of the canonical basis of $\mathbb{C}^n$).

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}^n_1$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \ldots, 0)$, of the canonical basis of $\mathbb{C}^n$). The map $U \mapsto Ue_1$ is a surjection of $\mathbb{U}(n)$ to $\mathbb{C}^n_1$, the unit sphere in $\mathbb{C}^n$.

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}_1^n$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \ldots, 0)$, of the canonical basis of $\mathbb{C}^n$). The map $U \mapsto Ue_1$ is a surjection of $\mathbb{U}(n)$ to $\mathbb{C}_1^n$, the unit sphere in $\mathbb{C}^n$. The inverse image of e_1 is those unitaries of the form $U = \mathbf{1} \oplus W$, under the direct sum decomposition $\mathbb{C}^n = [e_1] \oplus [e_1]^\perp$, where W is an arbitrary unitary on $[e_1]^\perp$.

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}_1^n$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \ldots, 0)$, of the canonical basis of $\mathbb{C}^n$). The map $U \mapsto Ue_1$ is a surjection of $\mathbb{U}(n)$ to $\mathbb{C}_1^n$, the unit sphere in $\mathbb{C}^n$. The inverse image of e_1 is those unitaries of the form $U = \mathbf{1} \oplus W$, under the direct sum decomposition $\mathbb{C}^n = [e_1] \oplus [e_1]^\perp$, where W is an arbitrary unitary on $[e_1]^\perp$. Thus the set of $W's$ is isomorphic to $\mathbb{U}(n-1)$ and, if $e_1 = \delta_1$, is canonically equal to it.

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}_1^n$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \ldots, 0)$, of the canonical basis of $\mathbb{C}^n$). The map $U \mapsto Ue_1$ is a surjection of $\mathbb{U}(n)$ to $\mathbb{C}_1^n$, the unit sphere in $\mathbb{C}^n$. The inverse image of e_1 is those unitaries of the form $U = \mathbf{1} \oplus W$, under the direct sum decomposition $\mathbb{C}^n = [e_1] \oplus [e_1]^\perp$, where W is an arbitrary unitary on $[e_1]^\perp$. Thus the set of $W's$ is isomorphic to $\mathbb{U}(n-1)$ and, if $e_1 = \delta_1$, is canonically equal to it. This shows that the quotient group $\mathbb{U}(n)/\mathbb{U}(n-1)$ of left cosets of $U(n-1)$ is exactly $\mathbb{C}_1^n$.

Proof of the Killip–Nenciu Theorem

Returning to $\mathbb{U}(n)$, fix a unit vector $e_1 \in \mathbb{C}_1^n$ (it may be helpful to think of e_1 as the first vector, $\delta_1 = (1, 0, \dots, 0)$, of the canonical basis of $\mathbb{C}^n$). The map $U \mapsto Ue_1$ is a surjection of $\mathbb{U}(n)$ to $\mathbb{C}_1^n$, the unit sphere in $\mathbb{C}^n$. The inverse image of e_1 is those unitaries of the form $U = \mathbf{1} \oplus W$, under the direct sum decomposition $\mathbb{C}^n = [e_1] \oplus [e_1]^\perp$, where W is an arbitrary unitary on $[e_1]^\perp$. Thus the set of $W's$ is isomorphic to $\mathbb{U}(n-1)$ and, if $e_1 = \delta_1$, is canonically equal to it. This shows that the quotient group $\mathbb{U}(n)/\mathbb{U}(n-1)$ of left cosets of $U(n-1)$ is exactly $\mathbb{C}_1^n$. The invariant measure on $\mathbb{C}_1^n$ is clearly the usual rotation invariant measure.

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$,

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f.

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f. By the above noted fact about topological products, we cannot make this choice continuous in f, but one can make it measurable, indeed continuous on $\mathbb{C}_1^n \setminus \{\mathbb{C} \cdot e_1\}$, as follows.

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f. By the above noted fact about topological products, we cannot make this choice continuous in f, but one can make it measurable, indeed continuous on $\mathbb{C}_1^n \setminus \{\mathbb{C} \cdot e_1\}$, as follows. Suppose f is not colinear with e_1. Then e_1, f span a two dimensional subspace $\mathcal{H}_f$. We can pick another vector $e_2 \in \mathcal{H}_f$ orthonormal to e_1 specifying it uniquely by demanding that $\kappa \equiv \langle f, e_2 \rangle > 0$.

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f. By the above noted fact about topological products, we cannot make this choice continuous in f, but one can make it measurable, indeed continuous on $\mathbb{C}_1^n \setminus \{\mathbb{C} \cdot e_1\}$, as follows. Suppose f is not colinear with e_1. Then e_1, f span a two dimensional subspace $\mathcal{H}_f$. We can pick another vector $e_2 \in \mathcal{H}_f$ orthonormal to e_1 specifying it uniquely by demanding that $\kappa \equiv \langle f, e_2 \rangle > 0$.

We also define $\beta \equiv \overline{\langle f, e_1 \rangle}$

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f. By the above noted fact about topological products, we cannot make this choice continuous in f, but one can make it measurable, indeed continuous on $\mathbb{C}_1^n \setminus \{\mathbb{C} \cdot e_1\}$, as follows. Suppose f is not colinear with e_1. Then e_1, f span a two dimensional subspace $\mathcal{H}_f$. We can pick another vector $e_2 \in \mathcal{H}_f$ orthonormal to e_1 specifying it uniquely by demanding that $\kappa \equiv \langle f, e_2 \rangle > 0$.

We also define $\beta \equiv \overline{\langle f, e_1 \rangle}$ so that $\beta \in \mathbb{D}$ and $f = \bar{\beta} e_1 + \kappa e_2$.

Proof of the Killip–Nenciu Theorem

To realize $\mathbb{U}(n)$ as a product of $\mathbb{U}(n-1)$ and $\mathbb{C}_1^n$, we must pick, for each $f \in \mathbb{C}_1^n$, an element $\sigma(f) \in \mathbb{U}(n)$ so that $\sigma(f)e_1 = f$, i.e. $\sigma(f)$ is in the coset associated to f. By the above noted fact about topological products, we cannot make this choice continuous in f, but one can make it measurable, indeed continuous on $\mathbb{C}_1^n \setminus \{\mathbb{C} \cdot e_1\}$, as follows. Suppose f is not colinear with e_1. Then e_1, f span a two dimensional subspace $\mathcal{H}_f$. We can pick another vector $e_2 \in \mathcal{H}_f$ orthonormal to e_1 specifying it uniquely by demanding that $\kappa \equiv \langle f, e_2 \rangle > 0$.

We also define $\beta \equiv \overline{\langle f, e_1 \rangle}$ so that $\beta \in \mathbb{D}$ and $f = \bar{\beta} e_1 + \kappa e_2$. Since f is a unit vector $\kappa = \sqrt{1 - |\beta|^2}$.

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

$$\mathbf{1} - 2\langle e_1 - f, \cdot\rangle(e_1 - f)/\|e_1 - f\|^2$$

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

$$\mathbf{1} - 2\langle e_1 - f, \cdot\rangle(e_1 - f)/\|e_1 - f\|^2$$

To find its matrix form in the e_1, e_2 basis, we note that its first column must be $\begin{pmatrix} \bar{\beta} \\ \kappa \end{pmatrix}$ since it takes e_1 to f. Its second column is then determined by orthonormality and the desire to have determinant -1 (i.e. a reflection). Thus

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

$$\mathbf{1} - 2\langle e_1 - f, \cdot\rangle(e_1 - f)/\|e_1 - f\|^2$$

To find its matrix form in the e_1, e_2 basis, we note that its first column must be $\begin{pmatrix} \bar{\beta} \\ \kappa \end{pmatrix}$ since it takes e_1 to f. Its second column is then determined by orthonormality and the desire to have determinant -1 (i.e. a reflection). Thus

$$\Theta(\beta) = \begin{pmatrix} \bar{\beta} & \kappa \\ \kappa & -\beta \end{pmatrix}$$

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

$$\mathbf{1} - 2\langle e_1 - f, \cdot\rangle(e_1 - f)/\|e_1 - f\|^2$$

To find its matrix form in the e_1, e_2 basis, we note that its first column must be $\begin{pmatrix} \bar{\beta} \\ \kappa \end{pmatrix}$ since it takes e_1 to f. Its second column is then determined by orthonormality and the desire to have determinant -1 (i.e. a reflection). Thus

$$\Theta(\beta) = \begin{pmatrix} \bar{\beta} & \kappa \\ \kappa & -\beta \end{pmatrix}$$

We define the *Householder reflection*, $\sigma(f)$, on $\mathbb{C}^n$ to be $\Theta(\beta) \oplus \mathbf{1}_{n-2}$ under $\mathbb{C}^n = \mathcal{H}_f \oplus \mathcal{H}_f^\perp$ (where $\mathbf{1}_k$ is the size k identity matrix).

Proof of the Killip–Nenciu Theorem

There is an obvious unitary map on $\mathcal{H}_f$ that takes e_1 to f, namely reflection, $\Theta(\beta)$, in the line along $e_1 + f$, which is clearly

$$\mathbf{1} - 2\langle e_1 - f, \cdot\rangle(e_1 - f)/\|e_1 - f\|^2$$

To find its matrix form in the e_1, e_2 basis, we note that its first column must be $\begin{pmatrix} \bar{\beta} \\ \kappa \end{pmatrix}$ since it takes e_1 to f. Its second column is then determined by orthonormality and the desire to have determinant -1 (i.e. a reflection). Thus

$$\Theta(\beta) = \begin{pmatrix} \bar{\beta} & \kappa \\ \kappa & -\beta \end{pmatrix}$$

We define the *Householder reflection*, $\sigma(f)$, on $\mathbb{C}^n$ to be $\Theta(\beta) \oplus \mathbf{1}_{n-2}$ under $\mathbb{C}^n = \mathcal{H}_f \oplus \mathcal{H}_f^\perp$ (where $\mathbf{1}_k$ is the size k identity matrix). $\sigma(f)$ is now as an operator on $\mathbb{C}^n$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$. This maps $\mathbb{U}(n)$ Borel bijectively to $\mathbb{C}_1^n \times \mathbb{U}(n-1)$. By the Diaconis-Shahshahani Formula, under this map, Haar measure on $\mathbb{U}(n)$ is the product of the rotation invariant measure on $\mathbb{C}_1^n$ and Haar measure on $\mathbb{U}(n-1)$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$. This maps $\mathbb{U}(n)$ Borel bijectively to $\mathbb{C}_1^n \times \mathbb{U}(n-1)$. By the Diaconis-Shahshahani Formula, under this map, Haar measure on $\mathbb{U}(n)$ is the product of the rotation invariant measure on $\mathbb{C}_1^n$ and Haar measure on $\mathbb{U}(n-1)$.

α_0 is defined by $z\varphi_0 = \rho_1\varphi_1 + \bar{\alpha}_0\varphi_0$ (since $\varphi_0^* = \varphi_0$),

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$. This maps $\mathbb{U}(n)$ Borel bijectively to $\mathbb{C}_1^n \times \mathbb{U}(n-1)$. By the Diaconis-Shahshahani Formula, under this map, Haar measure on $\mathbb{U}(n)$ is the product of the rotation invariant measure on $\mathbb{C}_1^n$ and Haar measure on $\mathbb{U}(n-1)$.

α_0 is defined by $z\varphi_0 = \rho_1\varphi_1 + \bar{\alpha}_0\varphi_0$ (since $\varphi_0^* = \varphi_0$), so $\langle z\varphi_0, \varphi_0 \rangle = \alpha_0$.

Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$. This maps $\mathbb{U}(n)$ Borel bijectively to $\mathbb{C}_1^n \times \mathbb{U}(n-1)$. By the Diaconis-Shahshahani Formula, under this map, Haar measure on $\mathbb{U}(n)$ is the product of the rotation invariant measure on $\mathbb{C}_1^n$ and Haar measure on $\mathbb{U}(n-1)$.

α_0 is defined by $z\varphi_0 = \rho_1\varphi_1 + \bar{\alpha}_0\varphi_0$ (since $\varphi_0^* = \varphi_0$), so $\langle z\varphi_0, \varphi_0 \rangle = \alpha_0$. In terms of U and the cyclic vector, e_1, $\alpha_0 = \langle f, e_1 \rangle = \beta$,

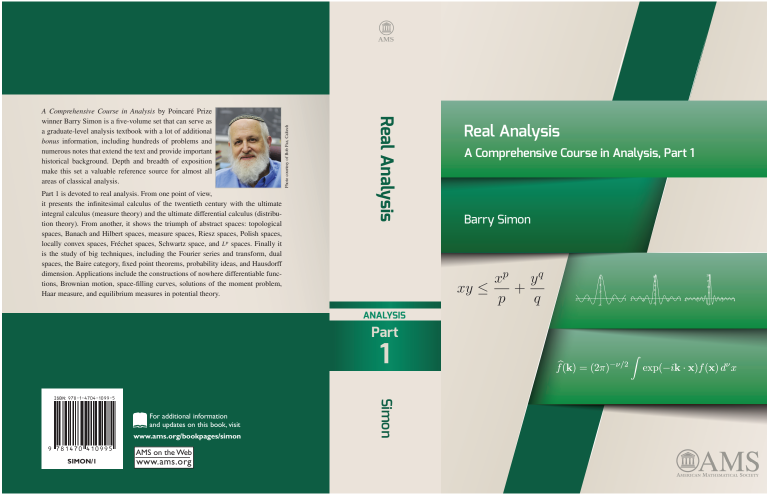
Proof of the Killip–Nenciu Theorem

Clearly $f \mapsto \sigma(f)$ is continuous on $\mathbb{C}^n \setminus \{\mathbb{C} \cdot e_1\}$ and discontinuous at the points of $\mathbb{C} \cdot e_1$. We define $\sigma(\lambda e_1) = \lambda \mathbf{1}$. We thus have shown that every unitary $U \in \mathbb{U}(n)$ can be uniquely written $\sigma(f)W$ where $f = Ue_1$ and W is a unitary map on $[e_1]^\perp$. This maps $\mathbb{U}(n)$ Borel bijectively to $\mathbb{C}_1^n \times \mathbb{U}(n-1)$. By the Diaconis-Shahshahani Formula, under this map, Haar measure on $\mathbb{U}(n)$ is the product of the rotation invariant measure on $\mathbb{C}_1^n$ and Haar measure on $\mathbb{U}(n-1)$.

α_0 is defined by $z\varphi_0 = \rho_1\varphi_1 + \bar{\alpha}_0\varphi_0$ (since $\varphi_0^* = \varphi_0$), so $\langle z\varphi_0, \varphi_0 \rangle = \alpha_0$. In terms of U and the cyclic vector, e_1, $\alpha_0 = \langle f, e_1 \rangle = \beta$, so the Householder reflection is the $\Theta(\alpha_0) \oplus \mathbf{1}_{n-2}$ of the AGR factorization and we see that the α_0 distribution is exactly, $d\kappa_{n-2}$. Moroever the Verblunsky coefficients of $\mathbb{U}(n-1)$ unitary are just $\{\alpha_{j+1}\}_{j=0}^{n-2}$, so by induction, the Killip–Nenciu Theorem is proven.

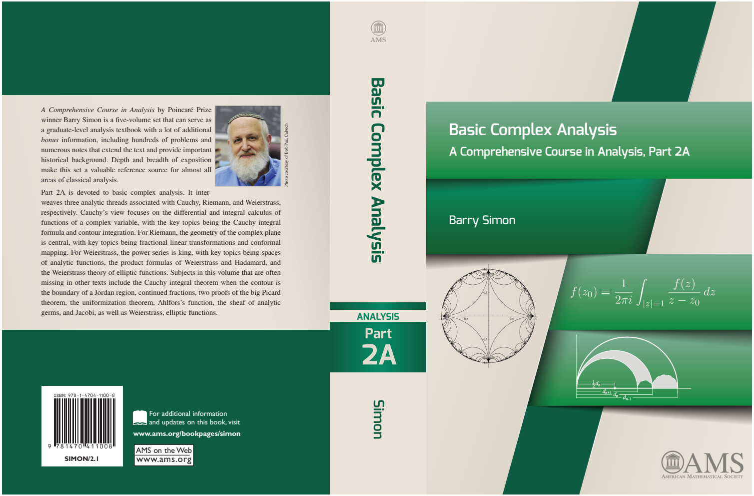
And Now a Word from Our Sponsor

And Now a Word from Our Sponsor



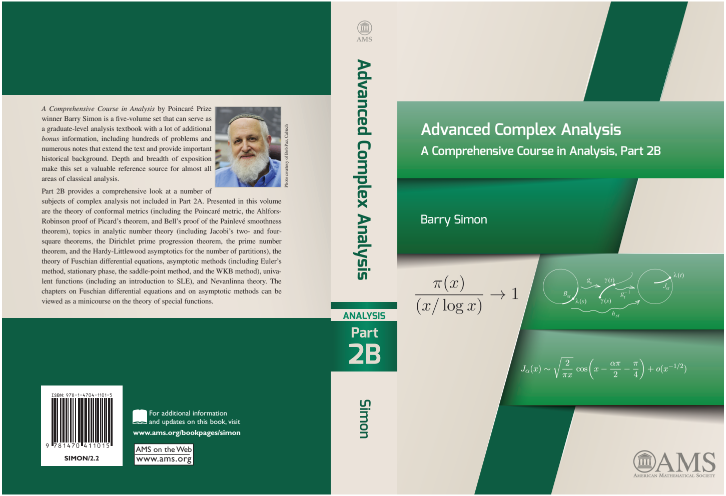
Google *simon comprehensive course*

And Now a Word from Our Sponsor



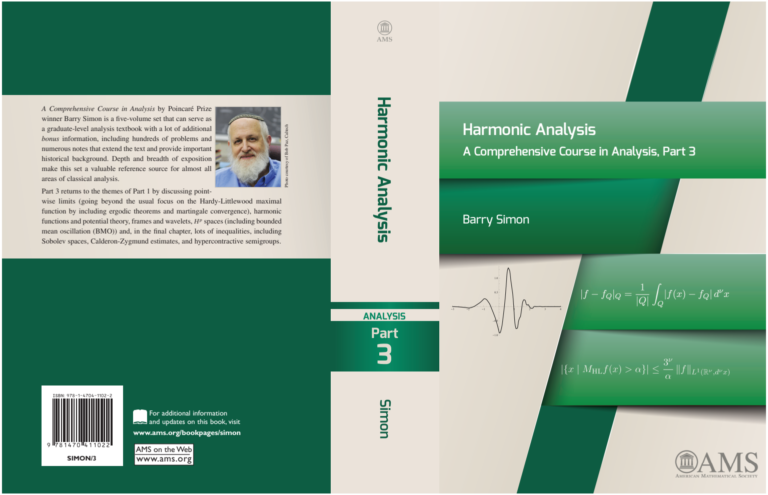
Google *simon comprehensive course*

And Now a Word from Our Sponsor



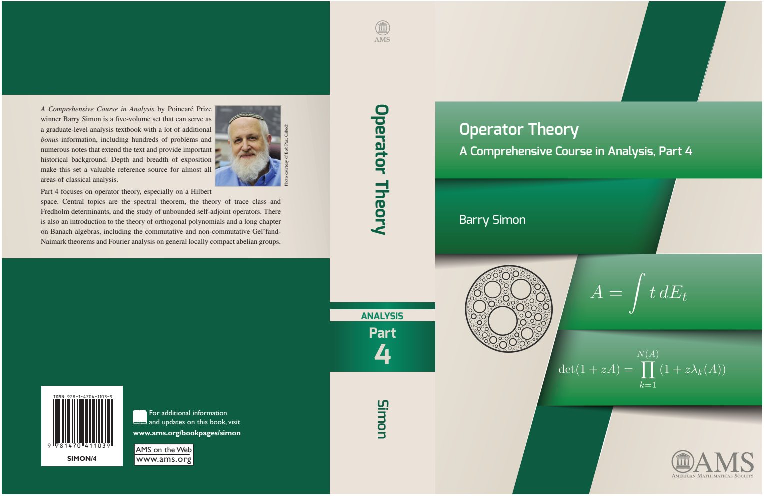
Google *simon comprehensive course*

And Now a Word from Our Sponsor



Google *simon comprehensive course*

And Now a Word from Our Sponsor



Google *simon comprehensive course*